

## **1. Introductory information**

### **o Title of the dataset**

Meteorological information from Romanian newspapers of the 19<sup>th</sup> century (ROMETNEWS19)

### **o For each file or group of similar files, a short description of what data it contains**

The file contains meteorological information between 1880 and 1900 extracted from three Romanian newspapers (România Liberă, Gazeta de Transilvania and Foaia Poporului). The information includes the date and location of identified meteorological phenomena, as well as impact and validation.

### **Contact information:**

Prof. dr. Sorin Cheval: [sorin.cheval@afahc.ro](mailto:sorin.cheval@afahc.ro); [sorincheval@yahoo.com](mailto:sorincheval@yahoo.com)

Dr. Aritina Haliuc: [aritina.haliuc@icub.unibuc.ro](mailto:aritina.haliuc@icub.unibuc.ro); [aritinahaliuc@gmail.com](mailto:aritinahaliuc@gmail.com)

## **2. Methodological information**

### **o Method description for collecting or generating the data, as well as the methods for processing data, if data other than raw data are being contributed**

Meteorological information was manually collected from the digital collections of three newspapers issued at the end of the 19<sup>th</sup> century in Romanian, (1) România Liberă (RL), issued at Bucharest (Romania), (2) Gazeta de Transilvania (GT), issued at Braşov, and (3) Foaia Poporului (FP), issued at Sibiu.

Each available newspaper issue was thoroughly examined and the meteorological events were identified and registered in a primary database with the following structure: 1) date of the meteorological event (both in Julian and Gregorian calendar); 2) location (name of the place, SIRUTA code, county); 3) event (e.g. rainfall, extreme high temperature, snowfall, thunderstorm); 4) category (e.g. storm, hail, flood, snowpack); 5) impacts (e.g. casualties, damages); 6) source (i.e. GT, RL, FP) and date of the publication YYYYMMDD (i.e. 18890620); 7) validation (i.e. information about the event are include in the records from the nearest weather station). The SIRUTA code was used to assign a correct and up-to-date location for each event. SIRUTA stands for Sistemul Informatic al Registrului Unităţilor Teritorial – Administrative (Informatics System of the Territorial – Administrative Units Register). SIRUTA is a classification used in Romania by the National Institute of Statistics (INS) to register Administrative-Territorial Units. Each unit has a numeric code updated every six months. The Territorial-Administrative Register is correlated with the Nomenclature of Territorial Statistics Units (NUTS) and it is structured on three levels: 1) counties and Bucharest municipality; 2) municipalities, towns, communes; and 3) localities, villages, and Bucharest sectors.

### **o Describe any quality-assurance procedures performed on the data**

Based on the results of the comparison between each newspaper report and meteorological records from the closest stations, three levels of validation were identified: (1) confirmed entries, for a match between newspaper and data from weather stations within 20 km-distance, at  $\pm 1$  day lag; (2) partially confirmed entries, for match between newspaper and data from weather stations within 20 km-distance, at  $\pm 2$  to 5-day lag, and (3) not confirmed entries, either because meteorological data are missing or they do not match with the newspaper event. The validation refers to a random sample of only 506 entries, from the total number of entries, and 142 entries (28.06%) of them were confirmed and partially confirmed.

### **3. Data specific information**

#### **o Full names and definitions (spell out abbreviated words) of column headings for tabular data**

Year (Julian calendar): year according to the Julian calendar

Month (Julian calendar): month according to the Julian calendar

Day (Julian calendar): day according to the Julian calendar

Year (Gregorian calendar): year according to the Gregorian calendar

Month (Gregorian calendar): month according to the Gregorian calendar

Day (Gregorian calendar): day according to the Gregorian calendar

Locality (as in the newspaper): name of the locality as in the newspaper report

Locality (current name): current name of the locality

SIRUTA Code: Each administrative unit in Romania has a numeric code updated every six months. The Territorial-Administrative Register is correlated with the Nomenclature of Territorial Statistics Units (NUTS) and it is structured on three levels: 1) counties and Bucharest municipality; 2) municipalities, towns, communes; and 3) localities, villages, and Bucharest sectors. Please, see here for more information: [http://colectaredate.insse.ro/senin/classifications.htm?selectedClassification=SIRUTA\\_S1\\_2018&action=general\\_information&classificationName=SIRUTA&classificationVersion=SIRUTA\\_S1\\_2018](http://colectaredate.insse.ro/senin/classifications.htm?selectedClassification=SIRUTA_S1_2018&action=general_information&classificationName=SIRUTA&classificationVersion=SIRUTA_S1_2018)

County: county where the locality is currently placed

Lat (Stereo 70): latitude coordinates Stereo 70 projection

Lon (Stereo 70): longitude coordinates Stereo 70 projection

Phenomenon (Romanian term): the name of the phenomenon described, in Romanian

Category (Romanian term): the category of the phenomenon described, in Romanian

Category (English term): the name of the category described, in English

Impact: information about the impact

Newspaper: FP – Foaia Poporului; GT – Gazeta de Transilvania; RL – România Liberă

Date of the issue (YYYYMMDD): YEAR (YYYY), MONTH (MM), DAY (DD) of the newspaper issue

Validation:

(1) confirmed entries, for a match between newspaper and data from weather stations within 20 km-distance, at  $\pm 1$  day lag; (2) partially confirmed entries, for match between newspaper and data from weather stations within 20 km-distance, at  $\pm 2$  to 5-day lag; (3) not confirmed entries, either because meteorological data are missing or they do not match with the newspaper event; (4) not checked; (5) data not available.

**o Units of measurement**

Not applicable, since the information are qualitative

**o Definitions for codes or symbols used to record missing data**

Not applicable

**o Specialized formats or abbreviations used**

Not applicable

**4. Sharing and Access information**